DEPARTMENT OF NATIONAL DEFENCE
CANADA

OPERATIONAL RESEARCH DIVISION

AORT RESEARCH NOTE RN-9703

EVALUATING FIGHTER LEAD-IN TRAINING (U)

by

G.L. CHRISTOPHER

Recommended by: _____ AORT TEAM LEADER

Approved by: _____ DOR(CAM)

Directorate Research Notes are written to document material which does not warrant or require more formal publication. The contents do not necessarily reflect the views of ORD or the Canadian Department of National Defence.

OTTAWA, CANADA APRIL 1997

ABSTRACT

Pilot candidates undergoing training to qualify to fly fighter aircraft in the Canadian Forces first proceed through a Fighter Lead-In Training (FLIT) course. After successful completion of the FLIT course, candidates enrol in a Basic Fighter Pilot (BFP) course. Recent changes to the FLIT training system did not permit all candidates to take the same FLIT course. Due to lack of time and capacity, fighter pilot candidates were divided into three groups and each group undertook a different FLIT course. All three groups were subsequently reunited to take the BFP course. Concerns over the possible impact of the three FLIT training streams on BFP course performance resulted in a study request to Air Operational Research. It was decided that the first action to be taken in the assessment of the impact of FLIT training would be to assess if BFP course performance was different for the three FLIT groups. Two statistical tests, the Kruskal-Wallis test and the one-way ANOVA General Linear Model procedure, were identified to evaluate the hypothesis that there are performance differences between the FLIT groups. The statistical tests were applied to mid-term BFP course results. The preliminary tests have shown that there is some evidence to support the contention that there are performance differences among the three FLIT groups. The results, however, will not be considered conclusive until the BFP course is completed.

RÉSUMÉ

Les candidats inscrits au programme de formation des pilotes de chasse des Forces armées canadiennes doivent d'abord réussir la formation de chef de patrouille, après quoi ils doivent suivre le cours élémentaire de pilote de chasse. Les changements récemment apportés à la formation de chef de patrouille ne permettaient pas à tous les candidats de suivre le même cours. En raison du manque de temps et de ressources, on a divisé les candidats en trois groupes qui ont chacun suivi un différent cours de formation de chef de patrouille. Les trois groupes ont ensuite été réunis pour suivre la formation élémentaire de pilote de chasse. Par suite de préoccupations exprimées sur les répercussions possibles de cette division en trois groupes sur les résultats du cours élémentaire de pilote de chasse, on a demandé à Recherche opérationnelle (Air) d'effectuer une étude. Il a été décidé que la première mesure à prendre pour évaluer l'incidence de la formation de chef de patrouille consisterait à déterminer s'il y avait des différences dans le rendement des trois groupes. Deux tests statistiques, soit le test de Kruskal-Wallis et l'uni-variée analyse de variance (ANOVA) modèle linéaire général, ont été retenus pour vérifier l'hypothèse d'un écart de rendement entre les trois groupes de candidats. Ces tests statistiques s'appliquent aux résultats à moyen terme du cours élémentaire de pilote de chasse. Les conclusions des tests préliminaires semblent confirmer qu'il y a une différence de rendement entre les trois groupes; cependant, ces conclusions ne seront pas considérées comme décisives avant la fin du cours élémentaire de pilote de chasse.

TABLE OF CONTENTS

ABSTRACT i

RÉSUMÉ. ii

TABLE OF CONTENTS iii

ACKNOWLEDGMENT iv

BACKGROUND 1

OBJECTIVE 2

MATHEMATICAL MODELS 3
Kruskal-Wallis Test 3
One-Way ANOVA 4

RESULTS 6
K-W Test 6
ANOVA Test 8

CONCLUSIONS 10

RECOMMENDATIONS 12

REFERENCES 14

ACKNOWLEDGMENT

The author wishes to acknowledge the contribution of Mr. Ed Emond of the Operational Research Division. Mr. Emond provided pertinent advice to assist in the identification of suitable statistical tests for this evaluation. Mr. Emond also applied the analysis of variance procedure to the mid-term results of the Basic Fighter Pilot course 9602.

EVALUATING FIGHTER LEAD-IN TRAINING

BACKGROUND

1. In the Canadian Forces fighter pilot training system, to qualify future pilots for the CF-18 fighter aircraft, candidates are trained and tested through a number of preparatory courses. Among the courses, candidates must succeed on a Fighter Lead-In Training (FLIT) course before going on to the Basic Fighter Pilot (BFP) course.

2. The normal routine would have pilot candidates undertake the FLIT course conducted by the Flying Instructors School (FIS), in Moose Jaw, Saskatchewan. Graduates from the FLIT course would then proceed to the next stage of their training to become fighter pilots, the BFP course. However, lack of time and training capacity required a group of pilot candidates, preparing to take the July 1996 BFP course, to be separated into three sub-groups. Each sub-group was then sent to a different FLIT course prior to attending the BFP course (Ref. 1).

3. The first sub-group, consisting of eight candidates, took the full FLIT course run by the FIS utilizing TUTOR (CT-114) aircraft. The second sub-group (six candidates) was trained using a modified version of the FIS FLIT course. The Modified FLIT course was essentially the FIS FLIT syllabus with the air-to-surface tactics and air combat manoeuvre training removed. The last sub-group, also with six candidates, was sent to a United States Air Force T-38 Conversion/IFF course run at Randolph Air Force Base, Texas.

4. These three FLIT sub-groups were united to undertake the BFP course commencing July 1996 (Serial 9602), conducted by 410 Squadron (Operational Training Unit) in Cold Lake, Alberta.

5. Before commencement of the BFP course, 410 Squadron raised concerns regarding the possible performance of the sub-groups, when brought together, as a result their different FLIT training backgrounds (Ref. 2). Included in

their concerns were:

a. that candidate performance on the BFP course would be related to the FLIT training background,

b. that some sub-group candidates would require extra training missions/flying hours to graduate,

c. that particular sub-groups might perform better or poorer in specific phases of the BFP training as a result of their FLIT training, and

d. the question of possible equipment solutions that could remedy FLIT training deficiencies and improve candidate performance at 410 Squadron.

OBJECTIVE

6. This study is the first step in the overall evaluation of Fighter Lead-In Training. This study was confined to be an assessment of the possibility that pilot candidate performance on the BFP course could be related to the type of FLIT training the candidates received. If performance differences between the sub-groups were found to exist, a follow-on investigation of remedies to correct these differences would be undertaken. Hence, the objective of this initial study was to assess whether there are performance differences between the FLIT sub-groups on the BFP course. The assessment would be conducted through the application of appropriate statistical tests on the candidates' course performance scores.

MATHEMATICAL MODELS

7. A literature review was initiated to identify suitable statistical tests that could be applied to the candidates' performance data. The statistical methods were to test the hypothesis that there was no difference between the BFP course performance levels of the three sub-groups. In other words, pilot candidate performance on the BFP course was unrelated to the FLIT training stream from which he or she came.

8. The literature review and consultations with statisticians in the Operational Research Division identified two statistical tests that would be appropriate to these circumstances. The first test accepted was the Kruskal-Wallis test, also referred to as the H Test (Ref. 3). The second statistical test selected was the one-way analysis of variance (one-way ANOVA) method (Refs. 3 and 4).

10. Kruskal-Wallis Test. The Kruskal-Wallis test is a non-parametric rank sum method. This model can be used to test the null hypothesis that k samples come from identical distributions. As a non-parametric model, it makes no assumptions about the nature of the sample data. For this reason the K-W method is very attractive as a test for differences in sample distributions. However, while the K-W test is very robust, it is less sensitive to small differences in sample distributions compared with other statistical tests.

9. To apply the K-W test, the data are ranked jointly, as though they are from one sample. Once the rank of each data element has been determined, the test statistic is calculated:

$$H = \frac{12}{n(n+1)} \cdot \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1) \qquad (1)$$

where: $n = n_1 + n_2 + \ldots + n_k$,

n_i is the number of data elements in the ith data group,

k is the number of data groups, and

R_i is the sum of the ranks for the ith data sample group.

10. If there are several data elements with the same value, the rank assigned to these data elements is the mean of the ranks they jointly occupy. For example, if the third, fourth and fifth ranked data elements have the same value, the three data elements would be assigned the rank of four (mean of three, four and five). The next data element in rank order would be given the rank of six, and the ranking would continue.

11. From sampling theory, the sampling distribution of **H** can be approximated by a chi-square distribution with $k - 1$ degrees of freedom (Ref. 4). For a given level of confidence, if the **H** value exceeds the chi-square distribution value, the null hypothesis would be rejected. Otherwise, the null hypothesis is accepted and the performance of the sample groups would be assumed to be the same.

12. <u>One-Way ANOVA</u>. One-way analysis of variance is a statistical method for comparing the means of several data sample groups. Whereas the K-W test did not make any assumptions about the data, one-way ANOVA assumes that the standard deviations of the sample groups are all equal. While this test is more sensitive to sample differences, one must have some confidence that the assumption of equal group standard deviations applies.

13. The model for one-way ANOVA is the General Linear Model. If samples are collected from k data sample groups, n_i would be the sample size for the ith group. Let x_{ij} represent the

jth observation from the ith sample group. The General Linear Model can then be represented by:

$$x_{ij} = \mu_i + \epsilon_{ij} \qquad (2)$$

where: μ_i is the mean of the distribution for sample group i, and

ϵ_{ij} is a random variation for observation x_{ij}.

The ϵ_{ij} are assumed to be simple random samples from a normal distribution with mean zero and standard deviation σ. The sample sizes n_i may differ, but the standard deviation σ is the same for all sample groups.

14. Before applying one-way ANOVA, one must have some confidence that the underlying assumption of equal sample standard deviation is valid. Reference 4 suggests that a simple general rule can be used to test the validity of the ANOVA assumption. The rule proposes that the ratio of the largest standard deviation of the sample groups to the smallest standard deviation must be less than two for the assumption to be taken as valid.

15. Many Statistical Analysis Software, SAS, applications have ANOVA implementations. Utilizing the SAS General Linear Model procedure on sample data from different groups produces two important values among the output components: the F statistic[1] and the probability that an F value greater than or equal to the F statistic could be produced by chance. The probability value can be interpreted as the likelihood that the data sample groups are the same, i.e. that the null hypothesis is true. For a detailed description of SAS General Linear Model procedure output, see Reference 3.

---

1. For an explanation of the F statistic and its application, see References 3 and 4.

RESULTS

16. Mid-term pilot candidate results from the BFP course were provided for preliminary evaluation. These results cover the first half of the Basic Fighter Pilot course. Each pilot candidate is rated on each component of the course. A scoring scheme is then applied by the course instructors to produce an overall score for each pilot candidate. The preliminary candidate scores for the first half of the BFP course are shown in Table I. For reasons of confidentiality, the names of the students will not be identified. The students will be referred to as candidate A, candidate B etc. The performance results, in terms of mean score and score range, for each FLIT group is shown in Figure 1.

17. K-W Test. Table II contains the results of the application of the Kruskal-Wallis test to the preliminary BFP course data. From Table II, it is seen that the rank sums (R_i) for the two FIS FLIT course streams are very similar. However, the Full FLIT group has two more students than the Modified FLIT group. The rank sum for the USAF T-38 group is approximately half the value for the other FLIT groups.

18. The resulting H test value is 3.72, while the Chi Square values for the 90 percent and 95 percent confidence levels[2] (10 and five percent probability of error) are 4.61 and 5.99, respectively. The K-W test for these levels of confidence would indicate that the null hypothesis should be accepted.

19. The K-W test suggests that there is not sufficient evidence to support the proposition that there is a performance difference among the three FLIT groups.

---

2. Confidence levels of 90 and 95 percent are accepted levels for statistical testing for evidence and strong evidence of rejection of the null hypothesis.

TABLE I: FIRST TERM BFP COURSE CANDIDATE PERFORMANCE SCORES

CANDIDATE	GROUP	SCORE		
Pilot A	Full FLIT	141		
Pilot B	Full FLIT	135		
Pilot C	Full FLIT	119		
Pilot D	Full FLIT	119	Mean =	119.1
Pilot E	Full FLIT	118	Std Dev =	13.09
Pilot F	Full FLIT	110		
Pilot G	Full FLIT	108		
Pilot H	Full FLIT	103		
Pilot I	Modified FLIT	124		
Pilot J	Modified FLIT	121		
Pilot K	Modified FLIT	112	Mean =	109.8
Pilot L	Modified FLIT	104	Std Dev =	11.11
Pilot M	Modified FLIT	102		
Pilot N	Modified FLIT	96		
Pilot O	USAF T-38	144		
Pilot P	USAF T-38	133		
Pilot Q	USAF T-38	133	Mean =	127.0
Pilot R	USAF T-38	119	Std Dev =	11.40
Pilot S	USAF T-38	118		
Pilot T	USAF T-38	15		

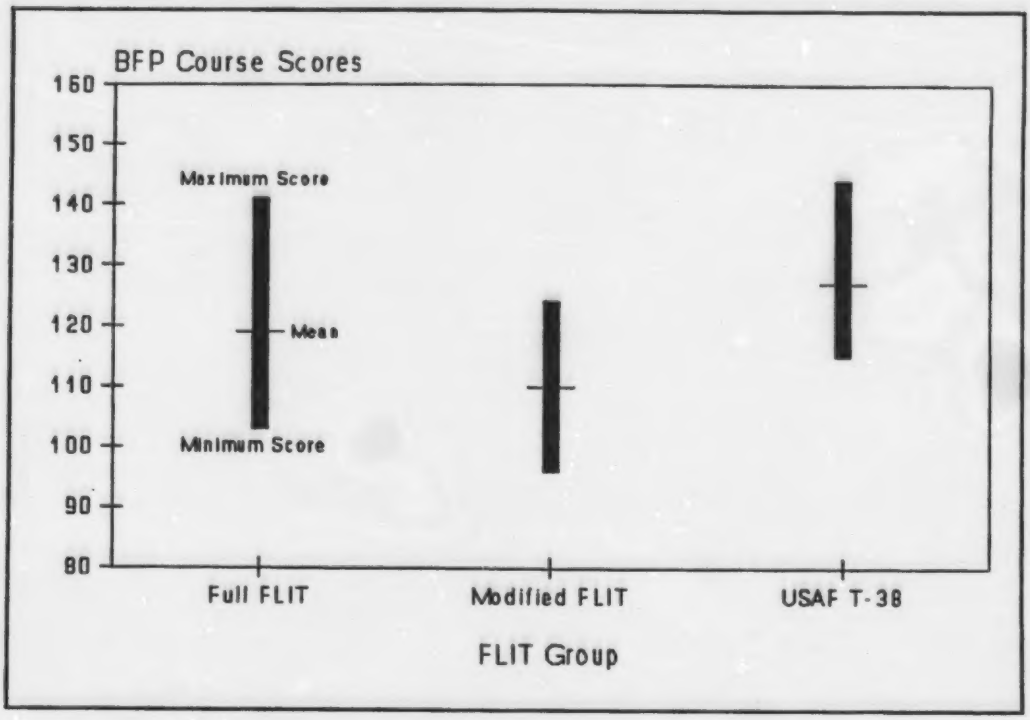

FIGURE 1 : BFP COURSE GROUP SCORES

20. ANOVA Test. The results of the ANOVA analysis of the preliminary performance scores of the candidates on the BFP course are shown in Table III.

21. From the Table, the greatest standard deviation for a group is 13.09 for the Full FLIT group, while the smallest standard deviation is 11.11 for the Modified FLIT group. The ratio of the largest standard deviation to the smallest is 1.18, well below the critical value of two, proposed as a general rule in Reference 3. The data pass the test for common standard deviation. The assumption required for the application of one-way ANOVA is considered valid and the ANOVA procedure can be applied.

TABLE II : KRUSKAL-WALLIS TEST RESULTS

Full FLIT Group	Rank	
Pilot A	2	
Pilot B	3	
Pilot C	9	
Pilot D	9	
Pilot E	11.5	
Pilot F	15	
Pilot G	16	Group Rank Sum
Pilot H	18	83.5
Mod FLIT Group	Rank	
Pilot I	6	
Pilot J	7	
Pilot K	14	
Pilot L	17	
Pilot M	19	Group Rank Sum
Pilot N	20	83
USAF T-38 Group	Rank	
Pilot O	1	
Pilot P	4.5	
Pilot Q	4.5	
Pilot R	9	
Pilot S	11.5	Group Rank Sum
Pilot T	13	43.5
H Value =	3.71	
Chi Square (.10,2) =	4.61	
Chi Square (.05,2) =	5.99	

TABLE III : ONE-WAY ANOVA ANALYSIS RESULTS

General Linear Model Procedure			
		Scores	
Group	Observations	Mean	Std Dev
Full FLIT	8	119.1	13.09
Modified Flit	6	109.8	11.11
USAF T-38	6	127.0	11.40
F Value =	3.06		
Probability (>F) =	0.074		

22. From Table III, the SAS General Linear Model procedure calculates a value for the F statistic of 3.06 and an associated probability of 0.074 (7.4 percent). This is to say that the probability that an F value of 3.06 or larger could occur by chance is 7.4 percent. Or alternatively, the probability that the distributions of group scores are the same and that the differences observed in the performance means are due to chance is 7.4 percent. Applying the accepted thresholds[2] for statistical significance, one can surmise that the ANOVA results indicate that there is some evidence to suggest that the null hypothesis of equal distributions should be rejected. According to the ANOVA analysis, there is some evidence that the course performance levels for the three FLIT groups are different.

CONCLUSIONS

23. The Air Operational Research Team was requested to examine the impact of three Fighter Lead-In Training courses on the performance of pilot candidates undertaking the Basic Fighter Pilot course. The first step in this investigation

was to determine if the performance of the pilot trainees on the BFP course was related to their FLIT training backgrounds. Through a literature review and consultation, two statistical tests were identified as appropriate to investigate the question of whether the performance of the pilot candidates on the Basic Fighter Pilot course is related to their training backgrounds. A non-parametric test, Kruskal-Wallis test, and one-way ANOVA were applied to the BFP course scores.

24. The Kruskal-Wallis test indicated that there was no evidence to reject the hypothesis that there are no performance differences among the three different FLIT groups. The ANOVA analysis showed that there is some evidence to support the assertion that the three FLIT groups are performing differently on the BFP course.

25. On first examination these results may appear contradictory. However, upon examination of the nature of these statistical tests, the results are logical. The K-W test is a very robust method because it makes no fundamental assumptions about the characteristics of the data samples to which it is being applied. Evidence to support a hypothe is provided by such a method is the ideal. However, the drawback to such non-parametric procedures is that they are often less sensitive to detection of small effects.

26. To detect small effects in data, more sensitive tests must be applied. To achieve the greater sensitivity, such statistical tests must make some assumptions about the nature of the data samples they are analysing. This is the case with the ANOVA general linear model procedure. This model assumes that all the data sample groups have the same standard deviation. Expecting that the distributions of the sample groups have this characteristic, permits smaller differences in sample means to be detected for a given sample size.

27. Given the characteristics of these two statistical testing procedures, it is not at all surprising that the ANOVA model might detect some evidence to support one hypothesis, while the Kruskal-Wallis test would not.

28. Given that some evidence exists to support the contention that there are performance differences among the three FLIT groups, an examination of Figure 1 would lead one to expect that the evidence is arising from the sample results of the Modified FLIT and USAF T-38 groups. If it is important to support this expectation, one could apply pair-wise tests to the group samples.

29. Given the small sample sizes, it is perhaps surprising that any evidence was found to suggest that there may be differences between the FLIT groups, even if differences did exist. From a statistical perspective, the evidence does not offer strong support. Before any actions are taken, it would be prudent to wait until the sample size was increased and stronger evidence for differences was found.

30. The main goal of this report was to describe and demonstrate statistical tests that are suitable to apply to the data in this study to assess the hypothesis that the different FLIT groups perform differently on the BFP course. Two such tests, the Kruskal-Wallis test and the one-way ANOVA General Linear Model procedure, were identified. Utilizing preliminary course data, the straightforward application of both tests was shown.

RECOMMENDATIONS

31. Group scores for the BFP course should be compiled when the course is complete, approximately June 1997. The two statistical tests should then be applied to the complete set of course data to determine if there is any evidence to indicate that there are overall performance differences among the three FLIT groups.

32. As well as testing for group differences in the overall course results, differences in group performance on the individual phases of the BFP course can also be checked.

33. Given anticipated individual performance variability among candidates within the same FLIT group, it can be expected that a larger sample size will be required to provide strong evidence of group performance differences. Six to eight data samples per group constitute a relatively small sample size. Before any actions are taken based on results from this BFP course, it may be prudent to confirm the conclusions by examining the results of a second BFP course involving candidates from the same FLIT streams.

34. Performance results from several past BFP courses should be compared with the latest course to assess whether there are overall changes occurring. Is average candidate performance changing? Is the variability in BFP course results increasing, decreasing, or stable? These issues should be investigated.

35. The practical significance of the group differences should also be considered. Even if the statistical tests indicate strong evidence that there are performance differences among the FLIT groups, the issue of whether the magnitudes of the differences warrant action should be considered. It is possible for group performance differences to be statistically significant, but the magnitudes of the differences are inconsequential in realistic terms. Some consideration should be given to specifying how large a performance difference is required before some action is taken. A cost-benefit assessment should be performed before any changes are implemented.

36. A final activity for this project should be an investigation and assessment of additional training aids (simulators and other technology) which could cost-effectively improve candidate preparation for, and performance on, the BFP course.

REFERENCES

1. Message SSO OT 031, 281446Z May 1996

2. Operational Research Division Project Request Form, Air Command Headquarters, SSO OT, 5 Oct 96

3. Moore, David S. and George P. McCabe, Introduction to the Practice of Statistics, W.H. Freeman and Company, 1989

4. Freund, John E. and Ronald E. Walpole, Mathematical Statistics-Fourth Edition, Prentice-Hall, 1987

UNCLASSIFIED
SECURITY CLASSIFICATION OF FORM
(Highest Classification of Title, Abstract, Keywords)

DOCUMENT CONTROL DATA
(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (the name and address of the organization preparing the document. Organizations for whom the document was prepared e.g. Establishment Sponsoring a contractor's report, or tasking agency, are entered in Section 8).

OPERATIONAL RESEARCH & ANALYSIS
DIRECTORATE OF AIR OPERATIONAL RESEARCH
NATIONAL DEFENCE HEADQUARTERS
OTTAWA, CANADA K1A 0K2

2. SECURITY CLASSIFICATION (overall security classification of the document, including special warning terms if applicable)

UNCLASSIFIED

3. TITLE (the complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title)

EVALUATING FIGHTER LEAD-IN TRAINING

4. AUTHORS (last name, first name, middle initial)

CHRISTOPHER, GARY, L.

5. DATE OF PUBLICATION (month Year of Publication of document)

APRIL 1997

6a. NO. OF PAGES (total containing information. Include Annexes, Appendices, etc.)
18

6b. NO. OF REFS (total cited in document)
4

7. DESCRIPTIVE NOTES (the category of document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

AORT RESEARCH NOTE

8. SPONSORING ACTIVITY (the name of the department project office or laboratory sponsoring the research and development. Include the address).

AIR COMMAND

9a. PROJECT OR GRANT NO. (if appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)

ACTIVITY 23212 TRAINING SYSTEMS

9b. CONTRACT NO. (if appropriate, the applicable number under which the document was written.)

10a. ORIGINATOR's document number (the official document number by which the document is identified by the originating activity. This number must be unique to this document.)

AORT RESEARCH NOTE RN 9703

10b. OTHER DOCUMENT NOS. (Any other numbers which may be assigned this document either by the originator or by the sponsor.)

11. DOCUMENT AVAILABILITY (any limitations on further dissemination of the document, other than those imposed by security classification.)

(X) Unlimited distribution
( ) Distribution limited to defence departments and defence contractors: further distribution only as approved
( ) Distribution limited to defence departments and Canadian defence contractors; further distribution only as approved
( ) Distribution limited to government departments and agencies; further distribution only as approved
( ) Distribution limited to defence departments; further distribution only as approved
( ) Other (please specify):

12. DOCUMENT ANNOUNCEMENT (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in 11) is possible, a wider announcement audience may be selected.)

13. ABSTRACT (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual).

Pilot candidates undergoing training to qualify to fly fighter aircraft in the Canadian Forces first proceed through a Fighter Lead-In Training (FLIT) course. After successful completion of the FLIT course, candidates enrol in a Basic Fighter Pilot (BFP) course. Recent changes to the FLIT training system did not permit all candidates to take the same FLIT course. Due to lack of time and capacity, fighter pilot candidates were divided into three groups and each group undertook a different FLIT course. All three groups were subsequently reunited to take the BFP course. Concerns over the possible impact of the three FLIT training streams on BFP course performance resulted in a study request to Air Operational Research. It was decided that the first action to be taken in the assessment of the impact of FLIT training would be to assess if BFP course performance was different for the three FLIT groups. Two statistical tests, the Kruskal-Wallis test and the one-way ANOVA General Linear Model procedure, were identified to evaluate the hypothesis that there are performance differences between the FLIT groups. The statistical tests were applied to mid-term BFP course results. The preliminary tests have shown that there is some evidence to support the contention that there are performance differences between the three FLIT groups. The results, however, will not be considered conclusive until the BFP course is completed.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus-identified . If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

PILOT TRAINING
FIGHTER LEAD-IN TRAINING
FIGHTER PILOT TRAINING
COURSE PERFORMANCE
STATISTICAL TESTS
SAMPLES
COMPARISON OF MEANS
COMPARISON OF DISTRIBUTIONS
H TEST
KRUSKAL-WALLIS TEST
RANK-SUM TEST
ANALYSIS OF VARIANCE
ANOVA